Identification of Matrix Joint Block Diagonalization

Yunfeng Cai, Ping Li
Cognitive Computing Lab
Baidu Research
No. 10 Xibeiwang East Road, Beijing 100193, China
10900 NE 8th St. Bellevue, Washington 98004, USA
{caiyunfeng, liping11}@baidu.com

Abstract

Given a set $\mathcal{C} = \{C_i\}_{i=1}^m$ of square matrices, the matrix blind joint block diagonalization problem (BJBDP) is to find a full column rank matrix A such that $C_i = A\Sigma_i A^\top$ for all i, where Σ_i's are all block diagonal matrices with as many diagonal blocks as possible. The BJBDP plays an important role in independent subspace analysis (ISA). This paper considers the identification problem for BJBDP, that is, under what conditions and by what means, we can identify the diagonalizer A and the block diagonal structure of Σ_i, especially when there is noise in C_i's. In this paper, we propose a "bi-block diagonalization" method to solve BJBDP, and establish sufficient conditions under which the method is able to accomplish the task. Numerical simulations validate our theoretical results. To the best of the authors' knowledge, existing numerical methods for BJBDP have no theoretical guarantees for the identification of the exact solution, whereas our method does.

1 Introduction

The matrix joint block diagonalization problem (JBDP) is a particular block term decomposition of a third order tensor (De Lathauwer, 2008; Nion, 2011). Over the past two decades, it has become a fundamental tool in independent subspace analysis (ISA) (e.g.,(Cardoso, 1998; Theis, 2006)). And ISA has found many applications in machine learning tasks, e.g., subspace clustering (Su et al., 2017; Wang et al., 2019; Ye et al., 2016), face recognition/verification (Cai et al., 2012; Le et al., 2011; Li et al., 2005, 2001), learning of disentangled representations (Awiszus et al., 2019; Stuehmer et al., 2020), etc. In this paper, we consider the identification problem for a blind JBDP. The results of this paper are naturally applicable to ISA. To be specific, next, we present the identification problem of the blind JBDP (BJBDP), then show how the problem arises in ISA.

Proceedings of the 24[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130.

1.1 Problem Statement

To introduce the identification problem of BJBDP, we need the following definitions.

Definition 1. We call $\tau_p = (p_1, \dots, p_\ell)$ a *partition* of positive integer p if $p_1, \dots, p_\ell$ are all positive integers and $\sum_{i=1}^{\ell} p_i = p$. The integer ℓ is called the *cardinality* of the partition τ_p, denoted by $\ell = \mathrm{card}(\tau_p)$. Two partitions $\tau_p = (p_1, \dots, p_\ell)$, $\tilde{\tau}_p = (\tilde{p}_1, \dots, \tilde{p}_{\tilde{\ell}})$ are said to be *equivalent*, denoted by $\tau_p \sim \tilde{\tau}_p$, if $\ell = \tilde{\ell}$ and there exists a permutation Π_ℓ such that $\tau_p = \tilde{\tau}_p \Pi_\ell$.

For example, $\tau_p = \{3,\ 1,\ 5,\ 2\}$, $\ell = \mathrm{card}(\tau_p) = 4$, $p = 11$, and $\tilde{\tau}_p = \{1,\ 5,\ 2,\ 3\}$ is equivalent to τ_p.

Definition 2. Given a partition $\tau_p = (p_1, \dots, p_\ell)$ and a matrix $X \in \mathbb{R}^{p\times p}$, partition X as $X = [X_{ij}]$ with $X_{ij} \in \mathbb{R}^{p_i \times p_j}$. Define the τ_p*-block diagonal part* and τ_p*-off-block diagonal part* of X, respectively, as

$$\mathrm{BlkDiag}_{\tau_p}(X) \triangleq \mathrm{diag}(X_{11}, \dots, X_{\ell\ell}),$$
$$\mathrm{OffBlkdiag}_{\tau_p}(X) \triangleq X - \mathrm{BlkDiag}_{\tau_p}(X).$$

The matrix X is referred to as *a* τ_p*-block diagonal matrix* if $\mathrm{OffBlkdiag}_{\tau_p}(X) = 0$.

$$X = \begin{array}{c} \\ p_1 \\ p_2 \\ \vdots \\ p_{\ell-1} \\ p_\ell \end{array} \begin{array}{c} \begin{array}{ccccc} p_1 & p_2 & \cdots & p_{\ell-1} & p_\ell \end{array} \\ \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1,\ell-1} & X_{1\ell} \\ X_{21} & X_{22} & \dots & X_{2,\ell-1} & X_{2\ell} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{\ell-1,1} & X_{\ell-1,1} & \dots & X_{\ell-1,\ell-1} & X_{\ell-1,\ell} \\ X_{\ell 1} & X_{\ell 2} & \dots & X_{\ell,\ell-1} & X_{\ell\ell} \end{bmatrix} \end{array},$$

$$\mathrm{BlkDiag}_{\tau_p}(X) = \begin{array}{c} \\ p_1 \\ p_2 \\ \vdots \\ p_{\ell-1} \\ p_\ell \end{array} \begin{array}{c} \begin{array}{ccccc} p_1 & p_2 & \cdots & p_{\ell-1} & p_\ell \end{array} \\ \begin{bmatrix} X_{11} & 0 & \dots & 0 & 0 \\ 0 & X_{22} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & X_{\ell-1,\ell-1} & 0 \\ 0 & 0 & \dots & 0 & X_{\ell\ell} \end{bmatrix} \end{array}.$$

The Joint Block Diagonalization Problem (JBDP) Given a matrix set $\mathcal{C} = \{C_i\}_{i=1}^m$ with $C_i \in \mathbb{R}^{d\times d}$ for $1 \le i \le m$. The JBDP for $\mathcal{C}$ with respect to a partition τ_p is to find a full column rank matrix $A = A(\tau_p) \in \mathbb{R}^{d\times p}$ such that all C_i's can be factorized as

$$C_i = A\Sigma_i A^\top = A\,\mathrm{diag}(\Sigma_i^{(11)}, \ldots, \Sigma_i^{(\ell\ell)})A^\top, \quad \forall i, \quad (1)$$

where Σ_i's are all τ_p-block diagonal. When (1) holds, we say that $\mathcal{C}$ is *τ_p-block diagonalizable* and A is a *τ_p-block diagonalizer* of $\mathcal{C}$.

The Blind JBDP (BJBDP) Given a matrix set $\mathcal{C} = \{C_i\}_{i=1}^m$ with $C_i \in \mathbb{R}^{d\times d}$ for $1 \le i \le m$. The BJBDP for $\mathcal{C}$ is to find a partition τ_p and a full column rank matrix $A = A(\tau_p)$ such that $\mathcal{C}$ is τ_p-block diagonalizable and $\mathrm{card}(\tau_p)$ is maximized. A solution to the BJBDP is denoted by (τ_p, A).

Uniqueness of BJBDP If (τ_p, A) with $\tau_p = (p_1, \ldots, p_\ell)$ and $A \in \mathbb{R}^{d\times p}$ is a solution to BJBDP, then $(\hat{\tau}_p, \widehat{A}) = (\tau_p\Pi_\ell, AD\Pi)$ is also a solution, where $\Pi_\ell \in \mathbb{R}^{\ell\times\ell}$ is a permutation matrix, D is any nonsingular τ_p-block diagonal matrix, $\Pi \in \mathbb{R}^{p\times p}$ is a permutation matrix associated with Π_ℓ, which permutes the column blocks of A as Π_ℓ permutes τ_p. In fact, Π can be obtained by replacing the 1 and 0 elements in the jth column of Π_ℓ by I_{p_j} and zero matrices of right sizes, respectively. If $(\hat{\tau}_p, \widehat{A}) = (\tau_p\Pi_\ell, AD\Pi)$, we say that (τ_p, A) and $(\hat{\tau}_p, \widehat{A})$ are *equivalent*, denoted by $(\tau_p, A) \sim (\hat{\tau}_p, \widehat{A})$. If any two solutions to BJBDP are equivalent, we say that the solution to the BJBDP is *unique*, the BJBDP for $\mathcal{C}$ is *uniquely τ_p-block-diagonalizable.*

Identifiability of BJBDP Let (τ_p, A) be a solution to the BJBDP for $\mathcal{C}$. Let $\widetilde{\mathcal{C}} = \{\widetilde{C}_i\}_{i=1}^m = \{C_i + E_i\}_{i=1}^m$, where $E_i \in \mathbb{R}^{d\times d}$ is a perturbation to C_i for $1 \le i \le m$. Under what conditions, and by what means, we can find a $(\tilde{\tau}_p, \widetilde{A})$ such that

$$\widetilde{C}_i \approx \widetilde{A}\widetilde{\Sigma}_i\widetilde{A}^\top = \widetilde{A}\,\mathrm{diag}(\widetilde{\Sigma}_i^{(11)}, \ldots, \widetilde{\Sigma}_i^{(\ell\ell)})\widetilde{A}^\top, \quad \forall i,$$

where $\widetilde{\Sigma}_i$'s are all $\tilde{\tau}_p$-block diagonal matrices with $\tilde{\tau}_p \sim \tau_p$, and $\widetilde{A}$ is close to A (up to block permutation and block diagonal scaling).

1.2 ISA: A Case Study

Independent Subspace Analysis (ISA) aims at separating linearly mixed unknown sources into statistically independent groups of signals. A basic model can be stated as

$$\mathbf{x} = A\mathbf{s},$$

where $\mathbf{x} \in \mathbb{R}^d$ is the observed mixture, $A \in \mathbb{R}^{d\times p}$ is the unknown mixing matrix and has full column rank, $\mathbf{s} \in \mathbb{R}^p$ is the source signal vector. Let $\mathbf{s} = \left[\mathbf{s}_1^\top, \ldots, \mathbf{s}_\ell^\top\right]^\top$ with $\mathbf{s}_j \in \mathbb{R}^{p_j}$ for $j = 1, \ldots, \ell$. Assume that each $\mathbf{s}_j$ has mean 0 and contains no lower-dimensional independent component, all $\mathbf{s}_j$ are independent of each other. ISA attempts to recover $\mathbf{s}$ from $\mathbf{x}$. Obviously, it holds that

$$C_{\mathbf{xx}} = \mathbb{E}(\mathbf{xx}^\top) = A\mathbb{E}(\mathbf{ss}^\top)A^\top = AC_{\mathbf{ss}}A^\top,$$

where $\mathbb{E}(\,\cdot\,)$ stands for expectation, and $C_{\mathbf{xx}}$, $C_{\mathbf{ss}}$ are the covariance matrices of $\mathbf{x}$ and $\mathbf{s}$, respectively. By assumption, $C_{\mathbf{ss}} = \mathrm{diag}(C_{\mathbf{s}_1\mathbf{s}_1}, \ldots, C_{\mathbf{s}_\ell\mathbf{s}_\ell})$ is τ_p-block diagonal, where $C_{\mathbf{s}_j\mathbf{s}_j}$ is the covariance matrix of $\mathbf{s}_j$.

Now let $\mathbf{x}(a), \ldots, \mathbf{x}(T)$ be T samples. In a piecewise stationary model (Lahat et al., 2012, 2014), the samples are partitioned into m non-overlapping domains $\{\mathcal{T}_i\}_{i=1}^m$, where $\mathcal{T}_i$ contains t_i samples, and $\sum_i t_i = T$. Let $\widetilde{C}_i \triangleq \frac{1}{t_i}\sum_{t\in\mathcal{T}_i}\mathbf{x}(t)\mathbf{x}(t)^\top$ and $\widetilde{\mathcal{C}} = \{\widetilde{C}_i\}_{i=1}^m$. Ideally, A is a τ_p-block diagonalizer of $\widetilde{\mathcal{C}}$. The question is that whether we can find $(\tilde{\tau}_p, \widetilde{A})$ by solving the BJBDP for $\widetilde{\mathcal{C}}$ such that $\tilde{\tau}_p \sim \tau_p$, and $\widetilde{A}$ is "close" to A? Under what conditions? And how?

1.3 A Short Review and Our Contribution

The identification problem is closely related to the uniqueness of the problem. In the context of ISA, it is shown that the decomposition of a random vector with existing covariance into independent, irreducible components is unique up to order and invertible transformations within the components (referred to as "trivial indeterminacy" hereafter) and an invertible transformation in possibly higher dimensional Gaussian component (Gutch and Theis, 2007, 2012). In the context of JBDP, when the matrices have additional structure, a local indeterminacy may occur (Gutch et al., 2010; Gutch and Theis, 2012). As JBDP is a particular block term decomposition of a third-order tensor, solution to JBDP is unique up to trivial determinacy almost surely (De Lathauwer, 2008).

Algorithmically, JBDP is usually formulated as an optimization problem, then solved via optimization-based numerical methods (e.g., (Cherrak et al., 2013; Nion, 2011)). However, without the information of the block diagonal structure, it is difficult to formulate the cost function. As a result, for BJBDP, a two-stage procedure is proposed – first apply a joint diagonalization method (Cardoso and Souloumiac, 1993; Ziehe et al., 2004), then reveal the block diagonal structure by certain clustering method (Tichavský et al., 2017). Such a procedure is based on a conjecture (Abed-Meraim and Belouchrani, 2004) that the JD and JBD problems share the same minima. But this conjecture is only partially proved (Theis, 2006). Three algebraic methods are proposed to solve BJBDP: When the diagonalizer is orthogonal, using matrix $*$-algebra, an error controlled method is proposed in Maehara and Murota (2011);

Using fundamental linear algebra, a recursive method is developed in Cai and Liu (2017) to compute the non-orthogonal diagonalizer, and quite recently, the method is non-trivially generalized to compute a blind block term decomposition of high order tensors (Cai and Li, 2021); Using the matrix polynomial, a three-stage method is proposed in Cai et al. (2019).

To the best of the authors' knowledge, current numerical methods for BJBDP have no theoretical guarantees for a good identification of the exact solution (τ_p, A), i.e., for the computed solution $(\tilde{\tau}_p, \widetilde{A})$, $\tilde{\tau}_p$ is equivalent to τ_p and $\widetilde{A}$ is "close" to A. In this paper, we will answer this fundamental question. For both noiseless and noisy cases, we first find the range space the diagonalizer via a (truncated) singular value decomposition (SVD) of a matrix, then reveal the block diagonal structure by a bi-diagonalization procedure. Under proper assumptions, we show that the proposed method is able to identify (τ_p, A). Numerical simulations validate our theoretical results.

The rest of this paper is organized as follows. In Section 2, we establish the identification condition of the range space of A and the block diagonal structure for both noiseless and noisy cases. Numerical experiments are presented in Section 3. Concluding remarks are given in Section 4.

Notation. I_n is the $n \times n$ identity matrix, and $0_{m\times n}$ is the m-by-n zero matrix. When their sizes are clear from the context, we may simply write I and 0. The symbol $\otimes$ denotes the Kronecker product. The operation $\mathrm{vec}(X)$ transforms a matrix X into a column vector formed by the first column of X followed by its second column and then its third column and so on. The spectral norm and Frobenius norm of a matrix are denoted by $\|\cdot\|_2$ and $\|\cdot\|_F$, respectively. For a matrix X, $\mathscr{R}(X)$ and $\mathscr{N}(X)$ stand for the range space and null space of X, respectively. For any square matrix set $\mathcal{D} = \{D_i\}_{i=1}^m$, we denote $\underline{D} = [D_1^\top, D_1, \ldots, D_m^\top, D_m]^\top$. For a subspace $\mathscr{V}$ of $\mathbb{R}^n$, its orthogonal complement is defined as $\mathscr{V}^\perp = \{w \in \mathbb{R}^n \mid w^\top v = 0, \forall v \in \mathscr{V}\}$.

2 Main Results

In this section, we establish the identification conditions for BJBDP. First, we identify $\mathscr{R}(A)$ in Section 2.1, then the block diagonal structure in Section 2.2

2.1 Identification of $\mathscr{R}(A)$

The following theorem identifies $\mathscr{R}(A)$ for the noiseless case, i.e., $E_i = 0$ for all $1 \le i \le m$.

Theorem 2.1. *Let (τ_p, A) be a solution to BJBDP for $\mathcal{C}$. Then $\mathscr{R}(A) = \mathscr{N}(\underline{C})^\perp = \mathscr{R}(\underline{C}^\top)$.*

By Theorem 2.1, it is natural for us to approximate of $\mathscr{R}(A)$ by the subspace spanned by the first p right singular vectors of $\underline{\widetilde{C}}$. The so called canonical angle is needed to state the result.

The canonical angles between two linear subspaces Let $\mathcal{X}, \mathcal{Y}$ be two k and ℓ dimensional subspaces of $\mathbb{R}^n$, respectively, and $k \ge \ell$. Let $X \in \mathbb{R}^{n\times k}, Y \in \mathbb{R}^{n\times \ell}$ be the orthonormal basis matrices of $\mathcal{X}$ and $\mathcal{Y}$, respectively, i.e.,

$$\mathcal{R}(X) = \mathcal{X},\ X^\top X = I_k, \quad \text{and} \quad \mathcal{R}(Y) = \mathcal{Y},\ Y^\top Y = I_\ell.$$

Denote ω_j for $1 \le j \le \ell$ the singular values of $Y^\top X$ in ascending order, i.e., $\omega_1 \le \cdots \le \omega_k$. The *canonical angles $\theta_j(\mathcal{X}, \mathcal{Y})$ between $\mathcal{X}$ and $\mathcal{Y}$* are defined by

$$0 \le \theta_j(\mathcal{X}, \mathcal{Y}) \triangleq \arccos \omega_j \le \frac{\pi}{2} \quad \text{for } 1 \le j \le \ell.$$

They are in descending order, i.e., $\theta_1(\mathcal{X}, \mathcal{Y}) \ge \cdots \ge \theta_\ell(\mathcal{X}, \mathcal{Y})$. Set

$$\Theta(\mathcal{X}, \mathcal{Y}) = \mathrm{diag}(\theta_1(\mathcal{X}, \mathcal{Y}), \ldots, \theta_\ell(\mathcal{X}, \mathcal{Y})).$$

It is worth mentioning here that the canonical angles defined above are independent of the choices of the orthonormal basis matrices X and Y.

Theorem 2.2. *Let (τ_p, A) be a solution to BJBDP for $\mathcal{C}$. Let the columns of V_2 be an orthonormal basis for $\mathscr{N}(A^\top)$, $\phi_1 \ge \cdots \ge \phi_d$ and $\tilde{\phi}_1 \ge \cdots \ge \tilde{\phi}_d$ be the singular values of $\underline{C}$ and $\underline{\widetilde{C}}$, respectively. Then*

$$\tilde{\phi}_p \ge \phi_p - \|\underline{E}\|, \qquad \tilde{\phi}_{p+1} \le \|\underline{E}\|. \tag{2}$$

In addition, let $\widetilde{U}_1 = [\tilde{u}_1, \ldots, \tilde{u}_p]$, $\widetilde{V}_1 = [\tilde{v}_1, \ldots, \tilde{v}_p]$, where $\tilde{u}_j$, $\tilde{v}_j$ are the left and right singular vector of $\underline{\widetilde{C}}$ corresponding to $\tilde{\phi}_j$, respectively, and $\widetilde{U}_1$, $\widetilde{V}_1$ are both orthonormal. If $\|\underline{E}\| < \frac{\phi_p}{2}$, then

$$\|\sin\Theta(\mathscr{R}(A), \mathscr{R}(\widetilde{V}_1))\| \le \frac{\|\widetilde{U}_1^\top \underline{E} V_2\|}{\tilde{\phi}_p}.$$

By Theorem 2.2, when $\|\underline{E}\|$ is sufficiently small compared with ϕ_p, we are able to find the correct p, and $\mathscr{R}(\widetilde{V}_1)$ is a good approximation for $\mathscr{R}(A)$.

2.2 Identification of the Block Diagonal Structure

In this section, we first discuss the identification of the block diagonal structure for the noiseless case, followed by the noisy case.

2.2.1 The Noiseless Case

This section is organized as follows:

(a) Firstly, we present a necessary and sufficient condition for when C_i's can be factorized in the form (1);

(b) Secondly, we present a way to determine whether the solution to the BJBDP is unique;

(c) Finally, we show how to find a solution to the BJBDP, and establish the theoretical guarantee.

Remark 1. The results for **(a)** and **(b)** are given below by Theorems 2.3 and 2.4, respectively. We need to emphasize here that Theorem 2.3 is rewritten from Cai and Liu (2017, Lemma 2.3), and Theorem 2.4 is partially rewritten from Cai and Liu (2017, Theorem 2.5). The difference between Theorem 2.3 and Lemma 2.3 is that the diagonalizer here is rectangular rather than square. The main difference between Theorem 2.4 and Theorem 2.5 is the proof. The proof here is simpler, more importantly, the proof is constructive and explainable. Borrowing those two results from Cai and Liu (2017) should not undermine the contribution of this paper, since they are the start point for our main contribution – the algorithms (Algorithms 2 and 4) to identify the solution of BJBDP with theoretical guarantees (Theorems 2.6 and 2.8).

The following linear space will play an important role in the analysis.

Definition 3. Given a matrix set $\mathcal{D} = \{D_i\}_{i=1}^m$ with $D_i \in \mathbb{R}^{q\times q}$, define

$$\mathscr{N}(\mathcal{D}) \triangleq \{X \in \mathbb{R}^{q\times q} \mid D_i X - X^\top D_i = 0,\ 1 \le i \le m\}.$$

Now we present a necessary and sufficient condition for when C_i's can be factorized in the form (1).

Theorem 2.3. *Given $\mathcal{C} = \{C_i\}_{i=1}^m$ with $C_i \in \mathbb{R}^{d\times d}$. Let $V_1 \in \mathbb{R}^{d\times p}$ be such that $V_1^\top V_1 = I_p$, $\mathscr{R}(V_1) = \mathscr{R}(\underline{C}^\top)$. Denote $B_i = V_1^\top C_i V_1$, $\mathcal{B} = \{B_i\}_{i=1}^m$. Then C_i's can be factorized as in (1) with $\mathscr{R}(A) = \mathscr{R}(\underline{C}^\top)$ if and only if there exists a matrix $X \in \mathscr{N}(\mathcal{B})$, which can be factorized into*

$$X = Y\,\mathrm{diag}(X_{11}, \ldots, X_{\ell\ell})Y^{-1}, \tag{3}$$

where $Y \in \mathbb{R}^{p\times p}$ is nonsingular, $X_{jj} \in \mathbb{R}^{p_j\times p_j}$ for $1 \le j \le \ell$ and $\lambda(X_{jj}) \cap \lambda(X_{kk}) = \emptyset$ for $j \ne k$.

According to Theorem 2.3, once we find an $X \in \mathscr{N}(\mathcal{B})$ which has a factorization in form (3), we can find a (τ_p, A) satisfying (1). Next, we examine some fundamental properties of $\Gamma \in \mathscr{N}(\{\Sigma_i\})$ with $\Sigma_i = \mathrm{diag}(\Sigma_i^{(11)}, \ldots, \Sigma_i^{(\ell\ell)})$, based on which we can determine whether (τ_p, A) is a solution to the BJBDP for $\mathcal{C}$.

Partition Γ as $\Gamma = [\Gamma_{jk}]$, where $\Gamma_{jk} \in \mathbb{R}^{p_j\times p_k}$. Using $\Sigma_i\Gamma - \Gamma^\top\Sigma_i = 0$, we have two sets of matrix equations. The first set is for $1 \le j = k \le \ell$:

$$\Sigma_i^{(jj)}\Gamma_{jj} - \Gamma_{jj}^\top\Sigma_i^{(jj)} = 0, \quad \text{for } 1 \le i \le m; \tag{4a}$$

The second set is for $1 \le j < k \le \ell$:

$$\begin{cases} \Sigma_i^{(jj)}\Gamma_{jk} - \Gamma_{kj}^\top\Sigma_i^{(kk)} = 0, \\ \Sigma_i^{(kk)}\Gamma_{kj} - \Gamma_{jk}^\top\Sigma_i^{(jj)} = 0, \end{cases} \quad \text{for } 1 \le i \le m. \tag{4b}$$

With the help of the Kronecker product, the first set of equations are equivalent to

$$G_{jj}\mathrm{vec}(\Gamma_{jj}) = 0, \tag{5a}$$

where

$$G_{jj} = \begin{bmatrix} I_{p_j} \otimes \Sigma_1^{(jj)} - [(\Sigma_1^{(jj)})^\top \otimes I_{p_j}]\Pi_j \\ \vdots \\ I_{p_j} \otimes \Sigma_m^{(jj)} - [(\Sigma_m^{(jj)})^\top \otimes I_{p_j}]\Pi_j \end{bmatrix},$$

$\Pi_j \in \mathbb{R}^{p_j^2\times p_j^2}$ is the perfect shuffle permutation matrix (Van Loan and Golub, 2012, Subsection 1.2.11) that enables $\Pi_j\mathrm{vec}(Z_{jj}^\top) = \mathrm{vec}(Z_{jj})$. The second set of equations are equivalent to

$$G_{jk}\begin{bmatrix} \mathrm{vec}(\Gamma_{jk}) \\ -\mathrm{vec}(\Gamma_{kj}^\top) \end{bmatrix} = 0, \tag{5b}$$

where

$$G_{jk} = \begin{bmatrix} I_{p_k} \otimes \Sigma_1^{(jj)} & (\Sigma_1^{(kk)})^\top \otimes I_{p_j} \\ I_{p_k} \otimes (\Sigma_1^{(jj)})^\top & \Sigma_1^{(kk)} \otimes I_{p_j} \\ \vdots & \vdots \\ I_{p_k} \otimes \Sigma_m^{(jj)} & (\Sigma_m^{(kk)})^\top \otimes I_{p_j} \\ I_{p_k} \otimes (\Sigma_m^{(jj)})^\top & \Sigma_m^{(kk)} \otimes I_{p_j} \end{bmatrix}.$$

For G_{jj} and G_{jk}, we introduce the following two properties:

(P1) For $1 \le j \le \ell$, for any $\mathrm{vec}(\Gamma_{jj}) \in \mathscr{N}(G_{jj})$, the eigenvalues of Γ_{jj} are the same real number or the same complex conjugate pair.

(P2) For $1 \le j < k \le \ell$, G_{jk} has full column rank.

The uniqueness of the solution to the BJBDP is closely related to **(P1)** and **(P2)**. In fact, we have the following theorem.

Theorem 2.4. *Let A be a τ_p-block diagonalizer of $\mathcal{C}$ i.e., (1) holds. Then (τ_p, A) is the unique solution to the BJBDP for $\mathcal{C}$ if and only if both* **(P1)** *and* **(P2)** *hold.*

Several important remarks follow in order.

Remark 2. Based on Theorem 2.4, once we get a τ_p-block diagonalizer A that factorizes C_i as (1), we can determine whether (τ_p, A) is the unique solution to the BJBDP by checking **(P1)** and **(P2)**.

Remark 3. By the proof of Theorem 2.4, we have the following facts which might help the understanding of **(P1)** and **(P2)**.

1) If **(P1)** does not hold for some j, then $\{\Gamma_i^{(jj)}\}_{i=1}^m$ can be further block diagonalized. This is because if **(P1)** does not hold for some j, there exists $\Gamma_{jj} \in \mathbb{R}^{p_j\times p_j}$ such that $\mathrm{vec}(\Gamma_{jj}) \in \mathscr{N}(G_{jj})$ and a nonsingular $W_j \in \mathbb{R}^{p_j\times p_j}$ such that

$$\Gamma_{jj} = W_j \,\mathrm{diag}(\Gamma_{jj}^{(a)}, \Gamma_{jj}^{(b)}) W_j^{-1}, \tag{6}$$

where $\Gamma_{jj}^{(a)}$ and $\Gamma_{jj}^{(b)}$ are two real matrices and $\lambda(\Gamma_{jj}^{(a)}) \cap \lambda(\Gamma_{jj}^{(b)}) = \emptyset$. Using $\mathrm{vec}(\Gamma_{jj}) \in \mathscr{N}(G_{jj})$, we have

$$\Sigma_i^{(jj)}\Gamma_{jj} - \Gamma_{jj}^\top \Sigma_i^{(jj)} = 0, \quad \text{for } 1 \le i \le m.$$

Substituting (6) into the above equality, we get

$$\widetilde{\Sigma}_i^{(jj)} \,\mathrm{diag}(\Gamma_{jj}^{(a)}, \Gamma_{jj}^{(b)}) - \mathrm{diag}(\Gamma_{jj}^{(a)}, \Gamma_{jj}^{(b)})^\top \widetilde{\Sigma}_i^{(jj)} = 0,$$

where $\widetilde{\Sigma}_i^{(jj)} = W_j^\top \Sigma_i^{(jj)} W_j$ for $i = 1, \ldots, m$. Partition $\widetilde{\Sigma}_i^{(jj)}$ as $\widetilde{\Sigma}_i^{(jj)} = \begin{bmatrix} \widetilde{\Sigma}_i^{(j11)} & \widetilde{\Sigma}_i^{(j12)} \\ \widetilde{\Sigma}_i^{(j21)} & \widetilde{\Sigma}_i^{(j22)} \end{bmatrix}$. Then it follows that

$$\begin{cases} \widetilde{\Sigma}_i^{(j12)}\Gamma_{jj}^{(b)} - (\Gamma_{jj}^{(a)})^\top \widetilde{\Sigma}_i^{(j12)} = 0, \\ \widetilde{\Sigma}_i^{(j21)}\Gamma_{jj}^{(a)} - (\Gamma_{jj}^{(b)})^\top \widetilde{\Sigma}_i^{(j21)} = 0, \end{cases} \quad \text{for } 1 \le i \le m.$$

Using $\lambda(\Gamma_{jj}^{(a)}) \cap \lambda(\Gamma_{jj}^{(b)}) = \emptyset$, we have $\widetilde{\Sigma}_i^{(j12)} = 0$ and $\widetilde{\Sigma}_i^{(j21)} = 0$. In other words, $\Sigma_i^{(jj)}$ for $1 \le i \le m$ can be further block diagonalized.

2) If **(P2)** does not hold for some $j \neq k$, then $\{\mathrm{diag}(\Sigma_i^{(jj)}, \Sigma_i^{(kk)})\}_{i=1}^m$ has a diagonalizer that is not (I_{p_j}, I_{p_k})-block diagonal. For example, let a_i's, b_i's and c_i's be arbitrary real numbers, it holds that

$$\begin{aligned} &\mathrm{diag}\left(\begin{bmatrix} 0 & a_i \\ a_i & b_i \end{bmatrix}, \begin{bmatrix} 0 & a_i \\ a_i & c_i \end{bmatrix} \right) \\ \equiv &\begin{bmatrix} 1&0&0&0\\0&1&-1&0\\0&0&1&0\\1&0&0&1 \end{bmatrix} \mathrm{diag}\left(\begin{bmatrix} 0 & a_i \\ a_i & b_i \end{bmatrix}, \begin{bmatrix} 0 & a_i \\ a_i & c_i \end{bmatrix} \right) \begin{bmatrix} 1&0&0&0\\0&1&-1&0\\0&0&1&0\\1&0&0&1 \end{bmatrix}^\top, \end{aligned}$$

in which the diagonalizer is not equivalent to I_4.

Remark 4. In the context of ISA, **(P1)** essentially requires the irreducibility (Gutch and Theis, 2007, 2012) of the independent component; **(P2)** generalizes the concept of "local indeterminacy/simple component" (Gutch et al., 2010; Gutch and Theis, 2012), and is much more mathematically strict.

Next, we consider how to solve the BJBDP.

Given a set $\mathcal{D} = \{D_i\}_{i=1}^m$ of q-by-q matrices with $\underline{D}$ having full column rank. When $\mathcal{D}$ has a $\tau_q = (q_1, q_2)$-block diagonalizer Z, i.e, D_i's can be factorized as $D_i = Z\Phi_i Z^\top$, where Φ_i's are τ_q-block diagonal, then set $X_* = Z^{-\top}\,\mathrm{diag}(I_{q_1}, -I_{q_2})Z^\top$ (Z is nonsingular since $\underline{D}$ has full column rank), it holds that

$$X_* \in \mathscr{N}(\mathcal{D}),\ (X_* - I)(X_* + I) = 0,\ X_* \neq \pm I.$$

Conversely, once we find such an X_*, factorize X_* into $X_* = Y\,\mathrm{diag}(I_{q_1}, -I_{q_2})Y^{-1}$, then $Y^{-\top}$ is a τ_q-block diagonalizer. In what follows, we formulate the problem of finding such an X_* as a constrained optimization problem.

Note that

$$\begin{aligned} &(X - I)(X + I) = 0 \\ \Leftrightarrow &\min_X \mathrm{tr}((X - I)^2(X + I)^2) \\ \Leftrightarrow &\min_X \mathrm{tr}(X^4) - 2\,\mathrm{tr}(X^2) + q, \end{aligned}$$

and $\mathrm{tr}(X) = 0$ together with $\mathrm{tr}(X^2) = q$ ensure $X \neq 0$ and the eigenvalues of X lie in both left and right complex plane, as a result, X is not a scalar matrix. So, we propose to find X_* by solving the following optimization problem:

$$\begin{aligned} \textsc{opt}(\mathcal{D}):\quad & \min_X \mathrm{tr}(X^4), \\ \text{subject to}\quad & X \in \mathscr{N}(\mathcal{D}), \mathrm{tr}(X) = 0, \mathrm{tr}(X^2) = q. \end{aligned} \tag{7}$$

For $\textsc{opt}(\mathcal{D})$, we have the following result.

Theorem 2.5. *Given a set $\mathcal{D} = \{D_i\}_{i=1}^m$ of q-by-q matrices with $\underline{D}$ having full column rank.*

(I) *If $\mathcal{D}$ does not have a nontrivial diagonalizer, then the feasible set of* $\textsc{opt}(\mathcal{D})$ *is empty.*

(II) *If $\mathcal{D}$ has a nontrivial diagonalizer, then* $\textsc{opt}(\mathcal{D})$ *has a solution X_*. In addition, assume*

$$\mu = \min_{\|z\|=1} \sqrt{\sum_{i=1}^m |z^{\mathrm{H}} D_i z|^2} > 0,$$

then X_* has two distinct real eigenvalues, and the gap between them are no less than two.

Remark 5. If $\underline{D}$ has full column rank, then $\mu > 0$ almost surely. Therefore, (2) holds almost surely without the assumption $\mu > 0$.

Based on Theorem 2.5, we present Algorithm 1, which will find a τ_q-diagonalizer Z for a matrix set $\mathcal{D} = \{D_i\}$ with $\mathrm{card}(\tau_q) = 2$ whenever $\mathcal{D}$ can be block-diagonalized.

Line 5 can be computed via Algorithm 7.6.3 in Van Loan and Golub (2012). The central task is to solve $\textsc{opt}(\mathcal{D})$. Using the Kronecker product,

Algorithm 1 Bi-Block Diagonalization (BI-BD)

Input: A set $\mathcal{D} = \{D_i\}_{i=1}^m$ of q-by-q matrices.
Output: (τ_q, Z) such that Z is a τ_q-block diagonalizer of $\mathcal{D}$ with $\tau_q = (q_1, q_2)$ or $\tau_q = (q)$.
if feasible set of OPT($\mathcal{D}$) is empty **then** set $\tau_q = (q)$, $Z = I_q$;
else Solve OPT($\mathcal{D}$), denote the solution by X_*;
  Compute $X_* = Y \operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1}$, where $\Gamma_1 \in \mathbb{R}^{q_1 \times q_1}$, $\Gamma_2 \in \mathbb{R}^{q_2 \times q_2}$, both $\lambda(\Gamma_1)$ and
  $\lambda(\Gamma_2)$ contain only one real number, and the two real numbers are different.
  Set $\tau_q = (q_1, q_2)$, $Z = Y^{-\top}$.
end if

$X \in \mathscr{N}(\mathcal{D})$ if and only if $\mathbf{L}(\mathcal{D})\operatorname{vec}(X) = 0$, where

$$\mathbf{L}(\mathcal{D}) \triangleq \begin{bmatrix} I_q \otimes D_1 - D_1^\top \otimes I_q \Pi_q \\ \vdots \\ I_q \otimes D_m - D_m^\top \otimes I_q \Pi_q \end{bmatrix} \in \mathbb{R}^{mq^2 \times q^2}. \quad (8)$$

Here $\Pi_q \in \mathbb{R}^{q^2 \times q^2}$ is the perfect shuffle permutation. The restarted Lanczos bi-diagonalization method (Baglama and Reichel, 2005) (MATLAB script `svds`), which is usually used to compute a few smallest/largest singular values and the corresponding singular vectors of a large scale matrix, is well suited here, since only the right singular vectors corresponding with the smallest singular value zero are needed. From the right singular vectors corresponding to zero, we can construct an orthonormal basis $\{X_1, \dots, X_s\}$ for $\mathscr{N}(\mathcal{D})$, where $s = \dim \mathscr{N}(\mathcal{D})$.

Now let $\mathcal{M} = [M_{ijkl}] \in \mathbb{R}^{s \times s \times s \times s}$, $K = [K_{ij}] \in \mathbb{R}^{s \times s}$ with $M_{ijkl} = \operatorname{tr}(X_i X_j X_k X_l)$, $K_{ij} = \operatorname{tr}(X_i X_j)$, the optimization problem OPT($\mathcal{D}$) is reduced into

$$\min_{\alpha \in \mathbb{R}^k} \mathcal{M}\alpha^4, \quad \text{subject to} \quad \alpha^\top K \alpha = 1, \quad (9)$$

where $\mathcal{M}\alpha^4 \triangleq \sum_{i,j,k,l} M_{ijkl}\alpha_i\alpha_j\alpha_l\alpha_l$. Let $K = G^\top G$ be the Cholesky factorization of K (by definition, K is symmetric positive definite), and denote $\beta = G\alpha$, $\mathcal{N} = \mathcal{M} \times_1 G^{-\top} \times_2 G^{-\top} \times_3 G^{-\top} \times_4 G^{-\top}$, where $\times_i$ denotes the modal product (Van Loan and Golub, 2012). Then (9) can be rewritten as

$$\min_{\beta \in \mathbb{R}^s} \mathcal{N}\beta^4, \quad \text{subject to} \quad \beta^\top \beta = 1, \quad (10)$$

whose KKT condition is $\mathcal{N}\beta^3 = \lambda\beta$, which is a Z-eigenvalue problem (Qi, 2005) of an order-4 tensor. Using the shifted power method (Cipolla et al., 2019; Kolda and Mayo, 2011), the eigenvector β_* corresponding with the smallest eigenvalue can be computed. Then X_* can obtained $X_* = \sum_{j=1}^s \alpha_j X_j$, where $\alpha = G^{-1}\beta_*$.

Algorithm 2 BJBDP via BI-BD

Input: A set $\mathcal{C} = \{C_i\}_{i=1}^m$ of d-by-d matrices.
Output: $(\hat{\tau}_p, \widehat{A})$, a solution to the BJBDP of $\mathcal{C}$.
Compute V_1, whose columns form an orthonormal basis for $\underline{C}^\top$;
Compute $\mathcal{B} = \{B_i\}_{i=1}^m = \{V_1^\top C_i V_1\}_{i=1}^m$;
Initialize $\hat{\tau}_p = (p)$, $\widehat{A} = V_1$, `list` $= [0]$;
while $\exists$ 0 in `list` **do**
  Find $t = \operatorname{argmax}\{\hat{\tau}_p(i) \mid \texttt{list}(i) = 0\}$;
  Set $k_1 = \sum_{i=1}^{t-1} \hat{\tau}_p(i) + 1$, $k_2 = \sum_{i=1}^{t} \hat{\tau}_p(i)$, $D_i = B_i(k_1 : k_2, k_1 : k_2)$ and $\mathcal{D} = \{D_i\}$;
  Call Algorithm 1 with input $\mathcal{D}$, denote the output by $(\hat{\tau}, \widehat{Z})$;
  if $\operatorname{card}(\hat{\tau}) = 1$ **then** Update `list`$(t) = 1$;
  else Update `list` and $\hat{\tau}_p$ by replacing their tth entry by $[0, 0]$ and $\hat{\tau}$, respectively;
    Update $B_i(k_1 : k_2, k_1 : k_2) = \widehat{Z}^{-1} D_i \widehat{Z}^{-\top}$, $\widehat{A}(:, k_1 : k_2) = \widehat{A}(:, k_1 : k_2)\widehat{Z}$.
  end if
end while

With the help of Algorithm 1, we may find a solution to BJBDP recursively. We summarize the method in Algorithm 2. Under proper assumptions, we can show that Algorithm 2 is able to identify the solution to BJBDP.

Theorem 2.6. *Assume that the* BJBDP *for* $\mathcal{C}$ *is uniquely* τ_p*-block-diagonalizable, and let* (τ_p, A) *be a solution satisfying* (1)*. Then* (τ_p, A) *can be identified via Algorithm* 2*, almost surely.*

2.2.2 The Noisy Case

In this section, we discuss the identification of the block diagonal structure with the presence of noise. According to Theorem 2.2, a good approximation for $\mathscr{R}(A)$ can be obtained when the perturbation is small. Given a perturbed matrix set $\widetilde{\mathcal{D}} = \{D_i + E_i\}_{i=1}^m$, where $\mathcal{D} = \{D_i\}_{i=1}^m$ can be block diagonalized, E_i is a perturbation to D_i. Inspired by the noiseless case, we consider an approximation of $\mathscr{N}(\mathcal{D})$ to approximately block-diagonalize $\widetilde{\mathcal{D}}$. The subspace $\mathscr{N}(\widetilde{\mathcal{D}})$ seems to be a natural choice, however, due to the presence of the noise, $\mathscr{N}(\widetilde{\mathcal{D}})$ in general only has a trivial element – the scalar matrix, which is useless for matrix joint block diagonalization. Recall (8), let $\tilde{v}_1, \dots, \tilde{v}_{p^2}$ be the right singular vectors of $\mathbf{L}(\widetilde{\mathcal{D}})$ corresponding to the singular values $\tilde{\sigma}_1, \dots, \tilde{\sigma}_{p^2}$, respectively, and the singular values be in a non-decreasing order. We define

$$\mathscr{N}_\delta(\widetilde{\mathcal{D}}) \triangleq \{\operatorname{reshape}(v, q, q) \mid v \in \mathscr{R}([\tilde{v}_1, \dots, \tilde{v}_k]), \\ \tilde{\sigma}_k \le \delta < \tilde{\sigma}_{k+1}\}.$$

Note that if $\delta = 0$ and $E_i = 0$ for all i, then $\mathscr{N}_\delta(\widetilde{\mathcal{D}}) = \mathscr{N}(\mathcal{D})$. So we may say that $\mathscr{N}_\delta(\widetilde{\mathcal{D}})$ is a generalization of $\mathscr{N}(\mathcal{D})$. In what follows, we will let $\mathscr{N}_\delta(\widetilde{\mathcal{D}})$ play the role of $\mathscr{N}(\mathcal{D})$. We also generalize the definition of diagonalizer as follows.

Definition 4. Given a set $\widetilde{\mathcal{D}} = \{\widetilde{D}_i\}_{i=1}^m$ of q-by-q matrices. We call Z *a* (τ_q, δ)*-diagonalizer* (also referred to as δ-diagonalizer) of $\widetilde{\mathcal{D}}$ if

$$\sum_{i=1}^m \|\widetilde{D}_i - Z\Phi_i Z^\top\|_F^2 \le C\,\delta^2,$$

where Φ_i's are all τ_q-block diagonal matrices, and C is a constant.

And rewrite the optimization problem $\text{OPT}(\mathcal{D})$ as

$$\begin{aligned} \text{OPT}(\widetilde{\mathcal{D}}, \delta):\quad & \min_X \operatorname{tr}(X^4), \\ \text{subject to}\quad & X \in \mathscr{N}_\delta(\widetilde{\mathcal{D}}), \operatorname{tr}(X) = 0, \operatorname{tr}(X^2) = q. \end{aligned}$$

Similar to Theorem 2.5, we have the next Theorem.

Theorem 2.7. *Given a set* $\widetilde{\mathcal{D}} = \{\widetilde{D}_i\}_{i=1}^m$ *of* q*-by-*q *matrices with* $\underset{\sim}{D}$ *having full column rank. Let* $\delta = o(1)$ *be a small real number.*

(I) *If* $\widetilde{\mathcal{D}}$ *does not have a nontrivial* δ*-diagonalizer, then the feasible set of* $\text{OPT}(\widetilde{\mathcal{D}}, \delta)$ *is empty.*

(II) *If* $\widetilde{\mathcal{D}}$ *has a nontrivial* δ*-diagonalizer, then* $\text{OPT}(\widetilde{\mathcal{D}}, \delta)$ *has a solution* X_**. In addition, assume*

$$\mu = \min_{\|z\|=1} \sqrt{\sum_{i=1}^m |z^{\mathrm{H}} \widetilde{D}_i z|^2} = O(1),$$

and for $i = 1, 2$*, let*

$$\mathrm{Rect}_i \triangleq \{z \in \mathbb{C} \,|\, |\mathrm{Re}(z) - \rho_i| \le a, |\mathrm{Im}(z)| \le b\},$$

where $a = O(\delta)$*,* $b = O(\delta)$*. Then*

$$\lambda(X_*) \subset \cup_{i=1}^2 \mathrm{Rect}_i, \quad \rho_1 - \rho_2 \ge 2 + O(\delta).$$

Based on Theorem 2.7, we have Algorithms 3 and 4. Specifically, Algorithm 3 finds a (τ_q, δ)-diagonalizer Z for a matrix set $\widetilde{\mathcal{D}} = \{\widetilde{D}_i\}$ with $\operatorname{card}(\tau_q) = 2$ whenever $\widetilde{\mathcal{D}}$ can be approximately block-diagonalized; Algorithm 4 finds an approximate solution to BJBDP with the presence of noise.

Finally, we establish the identifiability for BJBDP with the presence of noise. The modulus of non-divisibility and nonequivalence defined below are needed.

Definition 5. Let (τ_p, A) be a solution to BJBDP for $\mathcal{C}$ with $\mathrm{BlkDiag}_{\tau_p}(A^\top A) = I_p$. Let G_{jj}, G_{jk} be the same as in (5). The *modulus of irreducibility and nonequivalence* for $\mathcal{C}$ with respective to the diagonalizer A are respectively defined as

$$\omega_{\mathrm{ir}} \triangleq \begin{cases} \infty, & \tau_p = (1, \ldots, 1), \\ \min\limits_{p_j > 1} \{\sigma | \sigma \in \sigma(G_{jj}), \sigma \ne 0\}, & \text{otherwise}, \end{cases}$$

$$\omega_{\mathrm{neq}} \triangleq \omega_{\mathrm{neq}}(\mathcal{C}; A) = \min_{1 \le j < k \le t} \sigma_{\min}(G_{jk}).$$

Remark 6. The moduli ω_{ir} and ω_{neq} depend on the choice of the diagonalizer A. When the solution to BJBDP for $\mathcal{C}$ is unique, we can show that their dependency on diagonalizer A can be removed.

Algorithm 3 Approximate Bi-Block Diagonalization (A-BI-BD)

Input: A set $\widetilde{\mathcal{D}} = \{\widetilde{D}_i\}_{i=1}^m$ of q-by-q matrices, and a parameter δ.
Output: (τ_q, Z) such that Z is a (τ_q, δ)-block diagonalizer of $\widetilde{\mathcal{D}}$ with $\tau_q = (q_1, q_2)$ or $\tau_q = (q)$.
if feasible set of $\text{OPT}(\widetilde{\mathcal{D}}, \delta)$ is empty **then** set $\tau_q = (q)$, $Z = I_q$;
else Solve X_*;
  Compute $X_* = Y \operatorname{diag}(\Gamma_1, \Gamma_2) Y^{-1}$, where $\Gamma_1 \in \mathbb{R}^{q_1 \times q_1}$, $\Gamma_2 \in \mathbb{R}^{q_2 \times q_2}$, and the distance between $\lambda(\Gamma_1)$ and $\lambda(\Gamma_2)$ is approximately two.
  Set $\tau_q = (q_1, q_2)$, $Z = Y^{-\top}$.
end if

Algorithm 4 BJBDP via A-BI-BD

Input: A set $\widetilde{\mathcal{C}} = \{\widetilde{C}_i\}_{i=1}^m$ of d-by-d matrices.
Output: $(\hat{\tau}_p, \widehat{A})$ such that $\widehat{A}$ is a $(\hat{\tau}_p, \delta)$-block diagonalizer, where δ is a parameter.
Compute singular values $\tilde{\phi}_1 \ge \cdots \ge \tilde{\phi}_n$ and the corresponding right singular vectors $\tilde{v}_1, \ldots, \tilde{v}_n$ of $\underset{\sim}{C}$, set $\widetilde{V}_1 = [\tilde{v}_1, \ldots, \tilde{v}_p]$ with $\tilde{\phi}_{p+1} < \xi \tilde{\phi}_p$, where $\xi < 1$ is a real parameter, say $\xi = 0.1$;
Compute $\widetilde{\mathcal{B}} = \{\widetilde{B}_i\}_{i=1}^m = \{\widetilde{V}_1^\top \widetilde{C}_i \widetilde{V}_1\}_{i=1}^m$;
Initialize $\hat{\tau}_p = (p)$, $\widehat{A} = \widetilde{V}_1$, `list` $= [0]$;
while $\exists$ 0 in `list` **do**
  Find $t = \operatorname{argmax}\{\hat{\tau}_p(i) \mid \texttt{list}(i) = 0\}$;
  Set $k_1 = \sum_{i=1}^{t-1} \hat{\tau}_p(i) + 1$, $k_2 = \sum_{i=1}^{t} \hat{\tau}_p(i)$, $\widetilde{D}_i = \widetilde{B}_i(k_1 : k_2, k_1 : k_2)$ and $\widetilde{\mathcal{D}} = \{\widetilde{D}_i\}$;
  Call Algorithm 3 with input $\mathcal{D}$ and δ, denote the output by $(\hat{\tau}, \widehat{Z})$;
  if $\operatorname{card}(\hat{\tau}) = 1$ **then** Update `list`$(t) = 1$;
  else Update `list` and $\hat{\tau}_p$ by replacing their tth entry by $[0, 0]$ and $\hat{\tau}$, respectively; Update $B_i(k_1 : k_2, k_1 : k_2) = \widehat{Z}^{-1} D_i \widehat{Z}^{-\top}$, $\widehat{A}(:, k_1 : k_2) = \widehat{A}(:, k_1 : k_2)\widehat{Z}$.
  end if
end while

Remark 7. The modulus of irreducibility measures how far away the small blocks can be further block diagonalized; the modulus of nonequivalence measures how far away the BJBDP may have nonequivalent solutions.

The following theorem tells that when the noise is sufficiently small, (τ_p, A) can be identified.

Theorem 2.8. *Assume that the* BJBDP *for* $\mathcal{C} = \{C_i\}_{i=1}^m$ *is uniquely* τ_p*-block-diagonalizable, and let* (τ_p, A) *be a solution satisfying* (1). *Let* $\widetilde{\mathcal{C}} = \{\widetilde{C}_i\}_{i=1}^m = \{C_i + E_i\}_{i=1}^m$ *be a perturbed matrix set of* $\mathcal{C}$. *Denote*

$$\tau_p = (p_1, \dots, p_\ell), \qquad \hat{\tau}_p = (\hat{p}_1, \dots, \hat{p}_{\hat{\ell}}),$$
$$A = [A_1, \dots, A_\ell], \qquad \widehat{A} = [\widehat{A}_1, \dots, \widehat{A}_{\hat{\ell}}],$$

where $(\hat{\tau}_p, \widehat{A})$ *is the output of Algorithm 4. Assume* $\mathcal{N}(G_{jj}) = \mathscr{R}(\mathrm{vec}(I_{p_j}))$ *for all* j, *where* G_{jj} *is defined in* (5a). *Also assume that* p *is correctly identified in Line 3 of Algorithm 4. Let the singular values of* $\underline{\widetilde{C}}$ *be the same as in Theorem 2.2,*

$$\epsilon = \frac{\|\underline{E}\|}{\tilde{\phi}_p}, \quad r = \frac{\sqrt{2(d+2C)}\;\tilde{\phi}_p\,\epsilon}{\sigma^2_{\min}(A)(1-\epsilon^2)},$$
$$g_j = \frac{\sqrt{2j}}{(\hat{\ell}-1)\kappa\sqrt{p}} - \max\{\frac{\kappa}{\omega_{\mathrm{neq}}}, \frac{1}{\omega_{\mathrm{ir}}}\}r, \text{ for } j = 1, 2,$$

where C *and* κ *are two constants.*

(I) *If* $g_1 > 0$, *then* $\hat{\ell} = \ell$, *and there exists a permutation* $\{1', 2', \dots, \ell'\}$ *of* $\{1, 2, \dots, \ell\}$ *such that* $p_j = \hat{p}_{j'}$. *In order words,* $\hat{\tau}_p \sim \tau_p$.

(II) *Further assume* $g_2 > \frac{r}{\omega_{\mathrm{ir}}}$, *then there exists a* τ_p*-block diagonal matrix* D *such that*

$$\|[\widehat{A}_{1'}, \dots, \widehat{A}_{\ell'}] - AD\|_F$$
$$\le \frac{\frac{c\,r}{\omega_{\mathrm{neq}}}}{g_2 - \frac{r}{\omega_{\mathrm{ir}}}}\|A\|_F + (\frac{\epsilon^2}{\sqrt{1-\epsilon^2}} + \epsilon)\|\widehat{A}\|_F = O(\epsilon),$$

where c *is a constant.*

3 Numerical Experiment

In this section, we present several numerical examples. All numerical tests are carried out using MATLAB. Our method (BI-BD) is compared with two JBDP methods, namely, JBD-LM (Cherrak et al., 2013) and JBD-NCG (Nion, 2011), which are optimization based and need to know τ_p in advance.

Example 1. Given $\tau_p = (p_1, \dots, p_\ell)$, we generate the matrix set $\widetilde{\mathcal{C}} = \{\widetilde{C}_i\}_{i=1}^m$ as follows:

$$\widetilde{C}_i = AD_iA^\top + N_i, \quad i = 1, \dots, m,$$

where $A \in \mathbb{R}^{n\times p}$, $D_i \in \mathbb{R}^{p\times p}$ is τ_p-block diagonal and $N_i \in \mathbb{R}^{n\times n}$. The entries of A and D_i (block diagonal part) are drawn from $\mathcal{N}(0,1)$, and the entries of N_i from $\mathcal{N}(0,\sigma^2)$. The signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} = 10\log_{10} 1/\sigma^2$.

We carried out the tests with $m = 10$, $n = 15$, $p = 10$, $\tau_p = (2,3,3,4)$, $\mathrm{SNR} = 40, 60, 80, 100$. All tests are performed 20 times, and the average results are reported in Figures 1 to 4.

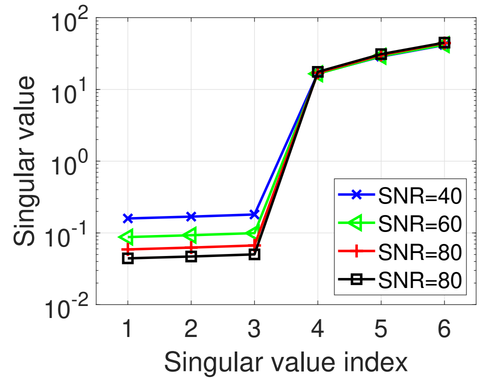

Figure 1: Singular values under different SNRs

Figure 1 plots the smallest six singular values of $\underline{\widetilde{C}}$ for different SNRs. We can see that there is a big gap between the third and fourth singular values. The larger SNR is, the larger the gap is. Therefore, we can find the correct p.

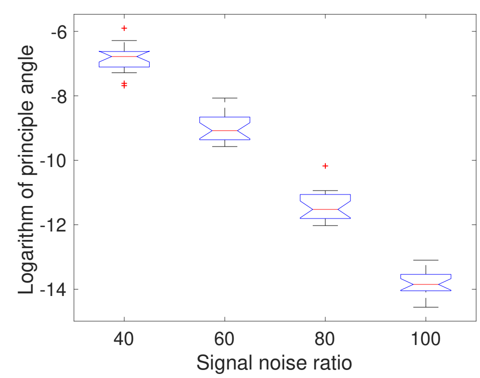

Figure 2: Principle angles under different SNRs

Figure 2 plots the principle angle between $\mathscr{R}(A)$ and the range space spanned by the right singular vectors of $\underline{\widetilde{C}}$ corresponding to the largest six singular vectors. We can see that $\mathscr{R}(A)$ is well estimated in all cases; the larger SNR is, the better the estimation is.

Figure 3 plots $\log(|\widehat{A}^\top A|)$, where $\widehat{A}$ is a diagonalizer obtained by BI-BD for SNR=40. We can see that the resulting matrix is approximately τ_p block diagonal up to permutation.

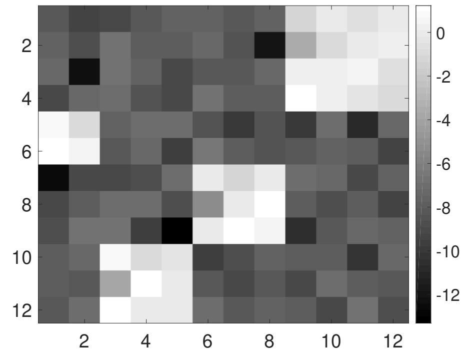

Figure 3: Structure of $|\widehat{A}^\top A|$

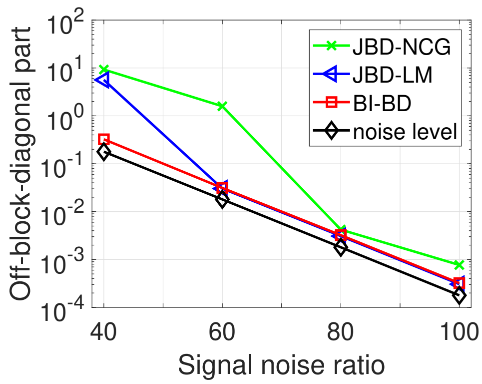

Figure 4: Norms of the off-block diagonal parts under different SNRs

Let $f(\widehat{A}) \triangleq \sqrt{\sum_{i=1}^m \| \mathrm{OffBlkdiag}_{\tau_p}(\widehat{A}^\dagger C_i \widehat{A}^{\dagger\top})\|_F^2}$. Figure 4 plots $f(\widehat{A})$ and the $p+1$st singular value $\tilde{\phi}_{p+1}$ of $\widetilde{C}$, where $\widehat{A}$ is the approximated diagonalizer, $\mathrm{BlkDiag}_{\tau_p}(\widehat{A}^\dagger \widehat{A}^{\dagger\top}) = I_p$, and $\widehat{A}^\dagger$ is the Moore–Penrose inverse. We can see that $f(\widehat{A})$ and $\tilde{\phi}_{p+1}$ decrease as SNR increases; BI-BD outperforms JBD-NCG and JBD-LM, especially when the SNR is small. In addition, $f(\widehat{A})$ corresponding with BI-BD is at the same order of $\tilde{\phi}_{p+1}$. Recall Theorem 2.2 that $\tilde{\phi}_{p+1}$ can be used as an estimation for the noise; Theorem 2.8 implies that $f(\widehat{A})$ should be at the order of the noise level. This explains why we observe $f(\widehat{A}) = O(\tilde{\phi}_{p+1})$.

Example 2. Consider three pieces of 3D independent sources. 6000 sample points were generated from noise free 3D wire-frames (as shown in the first row of Figure 5), then whitened. A random 9-by-9 matrix was used to mix the sources, and the mixed sources are shown in the second row of Figure 5. Our BI-BD method was applied to the mixed sources, and the recovered signals are shown in the last row of Figure 5. We can see that our method is able to recover the sources successfully.

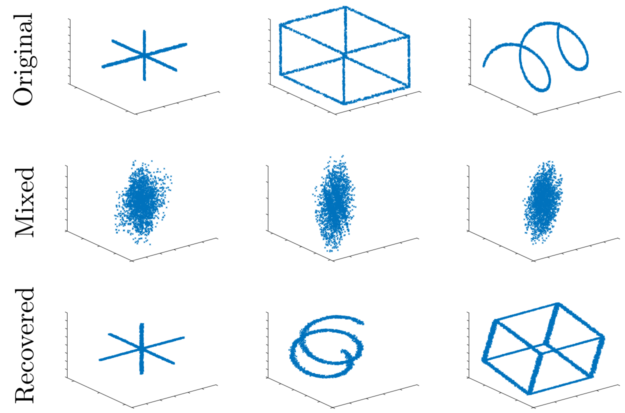

Figure 5: The original source signals, the mixed source signals, and the recovered signals

4 Conclusion

In this paper, we studied the identification problem for matrix joint block diagonalization. We propose a numerical method called BI-BD to solve the problem, in which the block diagonal structure is revealed step by step via solving an optimization problem. Under the assumption that the solution is unique, we show that BI-BD is able to identify the true solution when the noise is sufficiently small. Two parameters, namely, the modulus of irreducibility (which measures how far away the small blocks can be further block diagonalized) and the modulus of nonequivalence (which measures how far away the BJBDPmay have nonequivalent solutions), are introduced. According to Theorem 2.8, those two parameters determine the noise level that our BI-BD algorithm is able to identify the solution successfully. To the best of the authors' knowledge, our algorithm is the first method that has theoretical guarantees for the identification of the solution. Numerical simulations validate our theoretical results.

References

Karim Abed-Meraim and Adel Belouchrani. Algorithms for joint block diagonalization. In *Proceedings of the 12th European Signal Processing Conference*, pages 209–212, Vienna, Austria, 2004.

Maren Awiszus, Hanno Ackermann, and Bodo Rosenhahn. Learning disentangled representations via independent subspaces. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019.

James Baglama and Lothar Reichel. Augmented implicitly restarted lanczos bidiagonalization methods. *SIAM J. Sci. Comput.*, 27(1):19–42, 2005.

Xinyuan Cai, Chunheng Wang, Baihua Xiao, Xue Chen, and Ji Zhou. Deep nonlinear metric learning with independent subspace analysis for face verification. In *Proceedings of the 20th ACM Multimedia Conference (MM)*, pages 749–752, Nara, Japan, 2012.

Yunfeng Cai and Ping Li. A blind block term decomposition of high order tensors. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

Yunfeng Cai and Chengyu Liu. An algebraic approach to nonorthogonal general joint block diagonalization. *SIAM J. Matrix Anal. Appl.*, 38(1):50–71, 2017.

Yunfeng Cai, Guanghui Cheng, and Decai Shi. Solving the general joint block diagonalization problem via linearly independent eigenvectors of a matrix polynomial. *Numer. Linear Algebra Appl.*, 26(4):e2238, 2019.

Jean-François Cardoso. Multidimensional independent component analysis. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1941–1944, Seattle, WA, 1998.

Jean-François Cardoso and Antoine Souloumiac. Blind beamforming for non-gaussian signals. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 362–370. IET, 1993.

Omar Cherrak, Hicham Ghennioui, El Hossein Abarkan, and Nadège Thirion-Moreau. Non-unitary joint block diagonalization of matrices using a levenberg-marquardt algorithm. In *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, pages 1–5, Marrakech, Morocco, 2013.

Stefano Cipolla, Michela Redivo-Zaglia, and Francesco Tudisco. Shifted and extrapolated power methods for tensor ℓ^p -eigenpairs. *arXiv:1909.11964*, 2019.

Lieven De Lathauwer. Decompositions of a higher-order tensor in block terms-part II: Definitions and uniqueness. *SIAM J. Matrix Anal. Appl.*, 30(3): 1033–1066, 2008.

Harold W. Gutch and Fabian J. Theis. Independent subspace analysis is unique, given irreducibility. In *Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation (ICA)*, pages 49–56, London, UK, 2007.

Harold W Gutch and Fabian J Theis. Uniqueness of linear factorizations into independent subspaces. *Journal of Multivariate Analysis*, 112:48–62, 2012.

Harold W. Gutch, Takanori Maehara, and Fabian J. Theis. Second order subspace analysis and simple decompositions. In *Proceedings of the 9th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, pages 370–377, St. Malo, France, 2010.

W Kahan, BN Parlett, and Erxiong Jiang. Residual bounds on approximate eigensystems of nonnormal matrices. *SIAM J. Numer. Anal.*, 19(3):470–484, 1982.

Tamara G Kolda and Jackson R Mayo. Shifted power method for computing tensor eigenpairs. *SIAM J. Matrix Anal. Appl.*, 32(4):1095–1124, 2011.

Dana Lahat, Jean-François Cardoso, and Hagit Messer. Second-order multidimensional ica: Performance analysis. *IEEE Trans. Signal Process.*, 60(9):4598–4610, 2012.

Dana Lahat, Jean-Francois Cardoso, and Hagit Messer. Blind separation of multi-dimensional components via subspace decomposition: Performance analysis. *IEEE Trans. Signal Process.*, 62(11):2894–2905, 2014.

Quoc V. Le, Will Y. Zou, Serena Y. Yeung, and Andrew Y. Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3361–3368, Colorado Springs, CO, 2011.

Stan Z. Li, XiaoGuang Lv, and HongJiang Zhang. View-based clustering of object appearances based on independent subspace analysis. In *Proceedings of the Eighth International Conference On Computer Vision (ICCV)*, pages 295–300, Vancouver, Canada, 2001.

Stan Z Li, XiaoGuang Lu, Xinwen Hou, Xianhua Peng, and Qiansheng Cheng. Learning multiview face subspaces and facial pose estimation using independent component analysis. *IEEE Trans. Image Process.*, 14(6):705–712, 2005.

Takanori Maehara and Kazuo Murota. Algorithm for error-controlled simultaneous block-diagonalization of matrices. *SIAM J. Matrix Anal. Appl.*, 32(2): 605–620, 2011.

Dimitri Nion. A tensor framework for nonunitary joint block diagonalization. *IEEE Trans. Signal Process.*, 59(10):4585–4594, 2011.

Liqun Qi. Eigenvalues of a real supersymmetric tensor. *J. Symb. Comput.*, 40(6):1302–1324, 2005.

Gilbert W Stewart and Ji-Guang Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.

Jan Stuehmer, Richard E. Turner, and Sebastian Nowozin. Independent subspace analysis for unsupervised learning of disentangled representations. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1200–1210, Online [Palermo, Sicily, Italy], 2020.

Chunchen Su, Zongze Wu, Ming Yin, Kaixin Li, and Weijun Sun. Subspace clustering via independent subspace analysis network. In *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, pages 4217–4221, Beijing, China, 2017.

Ji-Guang Sun. On the variation of the spectrum of a normal matrix. *Linear Algebra Appl.*, 246:215–223, 1996.

Fabian J Theis. Towards a general independent subspace analysis. In *Advances in Neural Information Processing Systems*, pages 1361–1368, MIT Press, Cambridge, MA, 2006.

Petr Tichavský, Anh Huy Phan, and Andrzej Cichocki. Non-orthogonal tensor diagonalization. *Signal Process.*, 138:313–320, 2017.

Charles F Van Loan and Gene H Golub. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 4th edition, 2012.

Xing Wang, Jun Wang, Carlotta Domeniconi, Guoxian Yu, Guoqiang Xiao, and Maozu Guo. Multiple independent subspace clusterings. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI)*, pages 5353–5360, Honolulu, HI, 2019.

Wei Ye, Samuel Maurus, Nina Hubig, and Claudia Plant. Generalized independent subspace clustering. In *Proceedings the IEEE 16th International Conference on Data Mining (ICDM)*, pages 569–578, Barcelona, Spain, 2016.

Andreas Ziehe, Pavel Laskov, Guido Nolte, and Klaus-Robert Müller. A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *J. Mach. Learn. Res.*, 5:777–800, 2004.